

Clustering of Paraffin-Based Hybrid Rocket Fuels Combustion Data

A. Rüttgers · A. Petrarolo · M. Kobald

Received: date / Accepted: date

Abstract Clustering was applied to image data of hybrid rocket combustion tests for a better understanding of the complex flow phenomena. Novel techniques such as hybrid rockets that allow for cost reductions of space transport vehicles are of high importance in space flight. However, the combustion process in hybrid rocket engines is still a matter of ongoing research and not fully understood yet. Recently, combustion tests with different paraffin-based fuels have been performed at the German Aerospace Center (DLR). For a detailed analysis, the combustion process has been captured with a high-speed video camera, which leads to a huge amount of images for each test. In the end, a large dataset with a total number of 30 000 images for each combustion test has to be analyzed. In order to catch the essential flow structures, the combustion dataset was clustered with a K-means++ algorithm. Since the algorithm might converge to local optimal solutions, expensive repetitions have been performed in order to ensure that a global

solution is found in the end. Furthermore, a detailed analysis was performed to find an adequate clustering algorithm in the first place and to estimate the number of relevant clusters K in each experiment. As a result, valuable insights into the different combustion phases were obtained and a comparison of the quality of the combustion flame in the different tests could be made. In particular, depending on the fuel formulation and oxidizer mass flow, differences in the transients and flame brightness were found.

Keywords Hybrid rocket combustion · Optical investigations · K-means algorithm · Clustering · Machine learning

List of symbols

C_i	single cluster (-)
$I(x, y)$	grayscale pixel intensity (-)
J	objective function (squared error) (-)
K	number of clusters (-)
\mathbf{x}_j	data point j (a single image) (-)
d	problem dimension (resolution of \mathbf{x}_j) (-)
n	number of data points in single test (-)
$s(\mathbf{x}_j)$	silhouette value of the data point \mathbf{x}_j (-)
\bar{x}, \bar{y}	image barycenter coordinates (-)
$f(K)$	evaluation function to determine K (-)
α_k	weight factor in $f(K)$ (-)
μ_i	mean of cluster C_i (centroid) (-)

1 Introduction

In comparison to classical solid or liquid rocket engines, hybrid rocket engines have several advantages. First, since the propellants are stored in two different states

A. Rüttgers
German Aerospace Center (DLR), Simulation and Software
Technology, High-Performance Computing Department, Lin-
der Hoehe, 51147 Cologne, Germany
Tel.: +49 2203 601 3754
E-mail: alexander.ruettgers@dlr.de

A. Petrarolo
German Aerospace Center (DLR), Institute of Space Propul-
sion, Propellants Department, Im Langen Grund, 74239
Hardthausen, Germany
Tel.: +49 6298 28 728
E-mail: anna.petrarolo@dlr.de

M. Kobald
German Aerospace Center (DLR), Institute of Space Propul-
sion, Propellants Department, Im Langen Grund, 74239
Hardthausen, Germany
Tel.: +49 6298 28 622
E-mail: mario.kobald@dlr.de

of matter, hybrid motors are safer than solids. This also contributes to the reduction of the total costs of the engine. Moreover, they are characterized by controllable thrust, including shut off and restart capability. With respect to liquid engines, they are mechanically simpler and, consequently, cheaper (Karabeyoglu et al. 2011). Finally, their performance is in between those of solid and liquid engines. However, due to the diffusion limited combustion process typical for this kind of engines (the propellants are not pre-mixed, but need to gasify and mix with each other before being able to react), hybrid systems using conventional polymeric fuels are characterized by poor regression rate performance (resulting in low thrust level). In order to overcome this problem, the so-called liquefying hybrid rocket fuels, such as paraffin-based ones, can be used. These fast burning fuels are characterized by low viscosity and surface tension and they experience a different combustion mechanism with respect to conventional polymeric fuels (Karabeyoglu et al. 2001). During the combustion, instead of pyrolysing, they form a thin liquid layer on the fuel surface, which becomes instable due to the high-speed gas flow in the fuel port (Karabeyoglu et al. 2002). The liquid layer instabilities produce, in turn, droplets that entrain in the gas flow, thus working like a spray injection along the length of the motor. This causes an increase in the fuel burning area and, consequently, an increase in the regression rate of the fuel grain. Unfortunately, the entrainment phenomenon in hybrid rocket combustion process is still not fully understood yet and a matter of ongoing research.

For a better understanding of the combustion process, the burning tests have been recorded with a high-speed video camera that is able to capture 10 000 frames per second. Since a combustion test takes 3 seconds, each measurement results in 30 000 images that have to be analyzed. Therefore, it is important to group the images into subsets such that the essential flow structures and their length in time can be identified. In each combustion test, at least three different flow phases are expected, which correspond to the ignition phase, the steady combustion state and the extinction of the flame. Furthermore, turbulent or other irregular phases might occur. Even if these turbulent structures only exist within a short period of time, they might strongly affect the overall combustion behaviour. Therefore, it is necessary to perform a clustering of the dataset, which is able to group the data into separate flow phases and furthermore to detect strongly irregular combustion phenomena.

In the following, K-means++, an improved variant of K-means clustering (MacQueen 1967; Lloyd 1982), is applied to the combustion data. Since the number of

clusters K is not known in advance, a detailed analysis has been performed to estimate K in each test. As a result, it is shown that there is not a single optimal choice for K but that for each test different values should be considered. From our point of view, this results can be generalized to the clustering of experimental datasets in other areas of fluid dynamics.

The remainder of this article is organized as follows: first, a short introduction on K-means++ is given in Section 2. In Section 3 the experimental set-up that is used to obtain the combustion dataset is described. Finally in Section 4, the relevance of the algorithm for the current dataset is justified and a detailed analysis of the experimental data, based on the result of the clustering, is given. The analysis gives detailed insights into the combustion process and allows to identify optimal experimental configurations for future tests.

2 Mathematical Formulation

Various clustering algorithms exist in the literature that are able to subdivide a dataset of n observations (see Jain (2010) for a detailed survey) and the adequate choice of an algorithm strongly depends on the problem. In Section 4, the combustion dataset is clustered with a K-means++ algorithm (Arthur and Vassilvitskii 2007) that is based on the classical K-means algorithm (MacQueen 1967; Lloyd 1982). The choice of this specific algorithm is justified in Section 4.1. K-means++ and K-means only differ in the initialization of the cluster centroids. In both cases an objective function

$$J = \sum_{i=1}^K \sum_{\mathbf{x}_j \in C_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2 \quad (1)$$

is minimized iteratively with $\|\cdot\|$ as the Euclidean distance metric, $\boldsymbol{\mu}_i$ as the centroids of the K clusters C_i and $\mathbf{x}_j \in C_i$ as set of observations in C_i . In contrast to K-means, K-means++ reduces the probability that the algorithm terminates with local optimal findings for $\min(J)$ as final output. This is achieved by using an improved seeding algorithm. K-means++ takes only the first centroid from a uniform random distribution. Then, all further cluster centroids are taken randomly one after another from a weighted probability distribution that scales with the squared distance to the closest centroids that have already been chosen (see Arthur and Vassilvitskii (2007) for a more detailed description). In average, this leads to a more homogeneous distribution of the initial centroids. Furthermore, it can be shown that the objective function J of the initial centroids chosen by K-means++ differs from the objective function of the optimal clustering only by a factor of $\mathcal{O}(\log K)$ in

expectation. Nevertheless, all combustion tests in Section 4 are clustered with ten independent runs of K-means++ to avoid finding local optimal solutions.

Due to the tenfold repetition of the clustering algorithm, the computing time for each test is in the order of 1-2 days. A common approach in the literature to reduce the computing time of K-means type clustering is to apply the algorithm to random subsets of the full dataset in each iteration step. This approach is usually denoted as mini-batch K-means (Sculley 2010). However, this leads to a stochastic behavior of the algorithm and hampers its convergence. For this reason, the algorithm is applied to the full dataset at any time in the following sections.

A major drawback of K-means type clustering is that the number of clusters K is a parameter that has to be determined a priori. Although this input quantity can be roughly estimated from background knowledge in combustion theory, a threefold analysis to determine K was conducted in this study.

First, the data is clustered with different values for K in a range from 1 to 10. Using the different results for the minimized objective function, a detailed analysis is performed to find an adequate value for K . In the literature, various approaches are proposed to conduct this kind of analysis, such as the gap statistic by Tibshirani et al. (2001) or an evaluation function $f(K)$ according to Pham et al. (2005). In the following, we employ the approach by Pham et al. (2005) since it obtains similar findings as the gap statistic but is computationally less expensive. Using this approach, the evaluation function $f(K)$ is defined as

$$f(K) = \begin{cases} 1 & \text{if } K = 1 \\ \frac{J(K)}{\alpha_K J(K-1)}, & \text{if } J(K-1) \neq 0, \forall K > 1 \\ 1 & \text{if } J(K-1) = 0, \forall K > 1 \end{cases} \quad (2)$$

$$\alpha_K = \begin{cases} 1 - \frac{3}{4d} & \text{if } K = 2 \text{ and } d > 1 \\ \alpha_{K-1} + \frac{1-\alpha_{K-1}}{6} & \text{if } K > 2 \text{ and } d > 1 \end{cases} \quad (3)$$

where $J(K)$ is the optimal solution of the objective function for K clusters, d is the number of dimensions and α_K is a weight factor. It is intuitively clear that, for a given set of data points \mathbf{x}_j , $J(K) \leq J(K-1)$ for all $K \geq 2$. According to Pham et al. (2005) the drop from $J(K-1)$ to $J(K)$ is more pronounced in the case that an additional structure, i.e. an additional cluster, is resolved in the dataset. Furthermore, the authors investigate $J(K)$ in the case that the distribution of the data points is uniform, i.e. there is no clustering of points in the dataset. By analyzing these situations, they determined a heuristic weight factor α_K such that, for a uniform data distribution, $\alpha_K J(K-1) \approx J(K)$.

In their analysis, the authors denote $\alpha_K J(K-1)$ as estimated distortion of the data and $J(K)$ as real distortion of the data. The key idea is that the estimated distortion represents the expected decrease for J when the number of clusters is increased from $K-1$ to K but there is no real structure in the data. But if an actual structure has been resolved, the evaluation function $f(K)$ should be smaller than 1 as the drop is larger than expected. Consequently, this specific values of K should be considered in more detail by the user. In this application, the problem dimension d is in the order of $\mathcal{O}(10^5)$ since every pixel of the image is considered as an additional problem dimension. Since α_K converges to 1 in the limit $d \rightarrow \infty$, equation (2) simplifies to $f(K) \approx J(K)/J(K-1)$ in this specific application. Nevertheless, a small value of $f(K)$ is still an indicator for a choice of K that gives well-defined clusters.

Second, even though the problem dimension is high, the centroids $\boldsymbol{\mu}_i$ can be visualized and give information on the mean combustion in the corresponding clusters C_i . Therefore, it is essential to give a physical interpretation of the phenomena that are shown in each cluster center. It has to be noted, however, that the optimal number of clusters depends on the leading question. If only the main flow phases of the combustion are relevant, K can be chosen relatively small. On the other hand, if short-time turbulent structures are relevant, K has to be chosen larger. Therefore, the second approach leads to an ambiguity for K but ensures that the results are relevant in practice.

Third, the silhouette of the final clustering for each test is analyzed in Section 4.3. The silhouette technique allows to estimate the quality of a clustering (see Rousseeuw (1987) for a detailed description). The basic idea is that for each observation \mathbf{x}_j , a silhouette value $s(\mathbf{x}_j)$ in the range from -1 to 1 is computed. This value gives information on how well \mathbf{x}_j belongs to its cluster C_i with respect to the clustering outcome. A high value close to 1 represents a good fit, a value close to 0 indicates that the data point is close to the decision boundary of at least two clusters and a negative value close to -1 indicates a mismatch for that specific observation. The silhouette values are computed for all observations in the cluster and then sorted according to their size starting with the largest value. Consequently, a cluster C_i is well resolved if most observations have high silhouette values $s(\mathbf{x}_j)$. Furthermore, the full clustering is close to optimal if all clusters are well resolved. In more detail, a silhouette value $s(\mathbf{x}_j)$ for $\mathbf{x}_j \in C_i$ is

defined as

$$s(\mathbf{x}_j) = \begin{cases} 0, & \text{if } \text{dist}(C_i, \mathbf{x}_j) = \text{dist}(C \neq C_i, \mathbf{x}_j) = 0 \\ \frac{\text{dist}(C \neq C_i, \mathbf{x}_j) - \text{dist}(C_i, \mathbf{x}_j)}{\max(\text{dist}(C_i, \mathbf{x}_j), \text{dist}(C \neq C_i, \mathbf{x}_j))}, & \text{else.} \end{cases} \quad (4)$$

Here,

$$\text{dist}(C_i, \mathbf{x}_j) = \frac{1}{\#\{\mathbf{x}_k | \mathbf{x}_k \in C_i\}} \sum_{\mathbf{x}_k \in C_i} \text{dist}(\mathbf{x}_k, \mathbf{x}_j) \quad (5)$$

is the mean distance between \mathbf{x}_j and all other observations \mathbf{x}_k in its corresponding cluster C_i and

$$\text{dist}(C \neq C_i, \mathbf{x}_j) = \min_{m \neq i} \left(\frac{1}{\#\{\mathbf{x}_l | \mathbf{x}_l \in C_m\}} \sum_{\mathbf{x}_l \in C_m} \text{dist}(\mathbf{x}_l, \mathbf{x}_j) \right) \quad (6)$$

denotes the mean distance between \mathbf{x}_j and all observations \mathbf{x}_l in the cluster $C \neq C_i$ as the next-closest of the $K - 1$ neighbor clusters with respect to C_i .

3 Combustion Tests

The combustion tests were performed at the German Aerospace Center (DLR), Institute of Space Propulsion in Lampoldshausen, at the test complex M11. An already existing modular combustion chamber, used in the past to investigate the combustion behaviour of solid fuel ramjets (Ciezki et al. 2003), was adjusted and used for the test campaigns at atmospheric pressure. A side view of the whole combustion chamber set-up is shown in Fig. 1. The optically accessible combustion chamber is 450 mm long, 150 mm wide and 90 mm high. The flow straighteners with the pre-chamber have a total length of 450 mm and the post chamber is 150 mm long. The oxidizer main flow is entering the combustion chamber from the left, after having passed two flow straighteners. The mass flow rate is adjusted by a flow control valve and it is measured with a Coriolis flow meter. A high frequency static pressure sensor is mounted in the combustion chamber. Ignition is done via an oxygen/hydrogen torch igniter from the bottom of the chamber. A test sequence, reported in Table 1, is programmed before the test and is run automatically by the test bench control system. More details about the test bench and test settings are given in Kobald et al. (2015); Petrarolo and Kobald (2016); Petrarolo et al. (2018).

In the framework of this research, all tests were done at atmospheric pressure and with an oxidizer mass flow

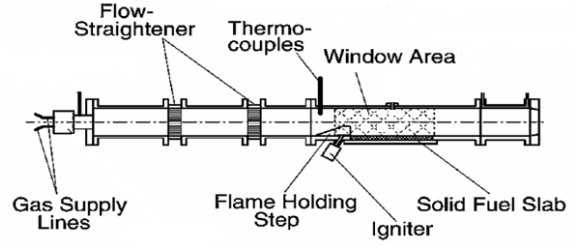


Fig. 1 Side view of the atmospheric combustion chamber set-up, adapted from Thumann and Ciezki (2002)

Table 1 Automatic test sequence

Time [s]	Action
T-0.2	Open ignition valves
T0	Open oxidizer main valve, start high-speed camera
T+0.3	Close ignition valves
T+3	Close oxidizer main valve, start nitrogen purge
T+5	End of sequence

ranging from 10 to 50 g/s. Combustion tests were performed using a single-slab paraffin-based fuel with a 20° forward facing ramp angle (see Fig. 2), in combination with gaseous oxygen. Two different fuel compositions were analyzed in this study: pure paraffin 6805 from the manufacturer Sasol Wax and the same paraffin with 5% mass addition of a commonly available polymer. All fuel slabs, produced and machined according to the same procedure, were 200 mm long, 100 mm wide and 20 mm high. Burning time was 3 seconds for each test. For video data acquisition a Photron Fastcam SA 1.1 high speed video camera was used with a maximum resolution of 1024x1024 pixel. The frame rate, resolution and shutter time of the camera were adjusted for each test, according to the test conditions and position of the camera. Tests were also performed using a CH* chemiluminescence imaging technique, with a band-pass filter centered around 431 nm placed in front of the camera. The excited CH* molecules emit photons around this wavelength, when they relax back to a lower energy state. Since high CH* concentration exists only in the main reaction zone, the resulting images provide a good indication of the instantaneous flame sheet location and topology.

In this study, four combustion tests have been analyzed. The test matrix is presented in Table 2.



Fig. 2 Fuel slab configuration used in this research, before (top) and after (bottom) combustion test

Table 2 Test matrix

Test no.	Fuel		$\dot{m}_{O_x} [g/s]$		CH* filter
	6805	6805+5% polymer	10	50	
284	✓			✓	✓
289		✓		✓	✓
296		✓	✓		✓
243		✓	✓		

4 Results and Discussion

This section is subdivided into three parts. First, a comparison of the outputs of different clustering algorithms on a low-dimensional approximation of the experimental dataset is performed, in order to justify the latter usage of a K-means++ algorithm. Then, the output of this clustering algorithm is analyzed to determine the number of clusters K in the full dataset. Finally, a physical interpretation of the different clusters and their centroids is given and the results are connected to the combustion tests.

4.1 Adequate choice of clustering algorithm

In the following, the choice of a K-means type clustering for this specific problem is justified. Since each observation $\mathbf{x}_j \in \mathbb{N}^d$ corresponds to a d -pixel grayscale image represented by integer values, the dimensionality d of the problem is high. For this reason, a comparison of the outcome of different clustering algorithms on a two- or three-dimensional representation of test 284 is performed since the test can be better visualized then. In Petrarolo and Kobald (2016), the experimental dataset is reduced in complexity with classical decomposition techniques such as a Principal Component Analysis and an Independent Component Analysis. In this study, an approximation based on image moments is used, as shown in Fig. 3. Here, the three-dimensional representation shows the mean image brightness μ of

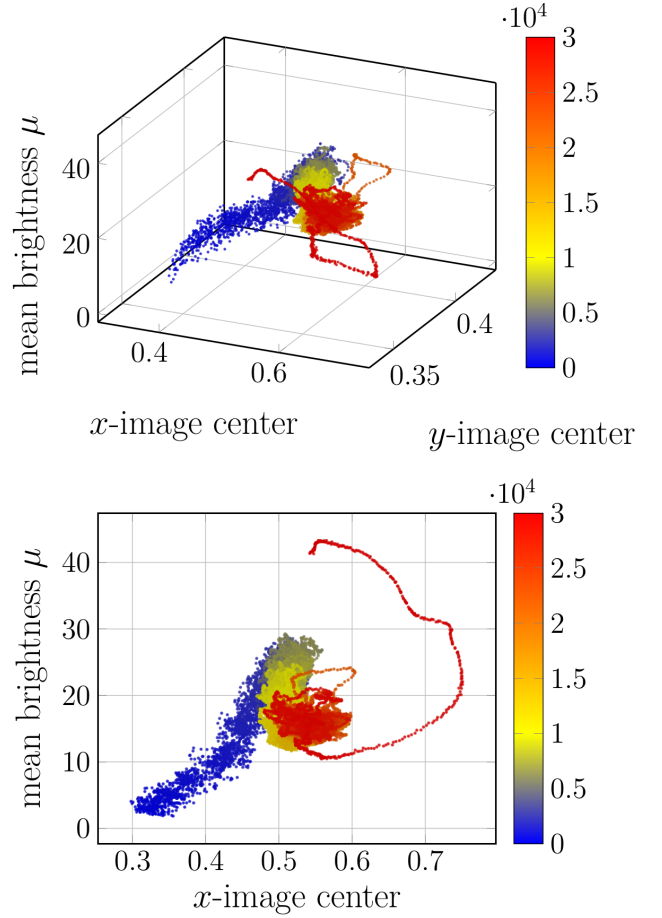


Fig. 3 Low-dimensional data approximation of test 284. The color specifies the index of each image in the dataset, i.e. the first few hundred images are colored blue, then the color changes to yellow and orange and the final frames with an index larger than 25000 are colored red.

all d -pixel images in the range from 0 to 255 and the x/y -coordinates of the corresponding image barycenters $\{\bar{x}, \bar{y}\}$. Let $I(x, y) \in \{0, 1, \dots, 255\}$ denote the grayscale pixel intensity at each pixel of the two-dimensional image, then the image barycenter is defined as

$$\{\bar{x}, \bar{y}\} = \left\{ \frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right\} \quad (7)$$

with

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y). \quad (8)$$

The values for the normalized image barycenters range from 0 to 1 in both coordinate directions. As an illustrative example, if \bar{x} is close to 1 and \bar{y} is close to 0, the brightest pixels of the image \mathbf{x}_j are in the lower right part of the corresponding image.

The upper part of Fig. 3 shows the triple $(\mu, \bar{x}, \bar{y})_j$ for all $j = 1, \dots, 30000$ images of test 284. Again, it has to be noted that this representation cannot capture all

relevant flow phenomena of the burning test but gives a general impression instead. For this reasons, an overlap of the data points in this representation does not necessarily mean that the corresponding images are identical. Furthermore, the colors in Fig. 3 represent the index j of each image and therefore show the temporal evolution of the combustion. First, it can be observed that the barycenters of the images vary only slightly in y -direction in the range $\approx 0.35 - 0.45$. For that reason, \bar{y} can be omitted without losing much relevant information and the focus is on the 2D subset $(\mu, \bar{x})_j$ in the following. Here, the first part of the experimental dataset colored in blue consists of low-brightness images with a combustion flame on the left side of the image. Later on, in Section 4.3, it will be shown that this part of the combustion corresponds to the first cluster C_1 . Next, the central part of the test with indices in the range $j \approx 5000 - 25000$ is colored in yellow and orange. In this case, the combustion flame is equally distributed over the images since $\bar{x}_j \approx 0.5$. Finally, the last phase of the test with indices $j > 25000$ is colored in red and consists of relative bright images \mathbf{x}_j where the combustion flame is located on the right side, i.e. $\bar{x}_j > 0.5$.

Although a low-dimensional representation of test 284 as in Fig. 3 gives only a coarse approximation of the real combustion test, essential requirements for an adequate clustering algorithm can already be identified. First, the data representation in Fig. 3 shows no isolated clusters but consists of a mixture of several high- and low-density data point regions. Due to the usage of a high-speed video camera, the difference of two consecutive images in brightness and position of the combustion flame is small and this kind of data representation has to be expected. For this reason, a clustering algorithm that can cope with not separated data points is required.

For a better understanding of the different possible approaches, a comparison of clustering algorithms provided by the Python machine learning library scikit-learn (Pedregosa et al. 2011) for the low-dimensional 2D representation of test 284 was performed. Fig. 4 shows the outcomes of clustering with K-means, spectral clustering (Ng et al. 2001), agglomerative clustering (Hastie et al. 2009) and DBSCAN (Ester et al. 1996) for this simplified problem. It is obvious that a density-based clustering such as DBSCAN is not optimal for the kind of problem under study. DBSCAN requires a hyperparameter ϵ as an input for the maximum distance of two observations that belong to the same cluster. If ϵ is chosen too large, only one cluster is found in the data of Fig. 4. On the other hand, if ϵ is chosen too small, a larger number of clusters is found in the same dataset.

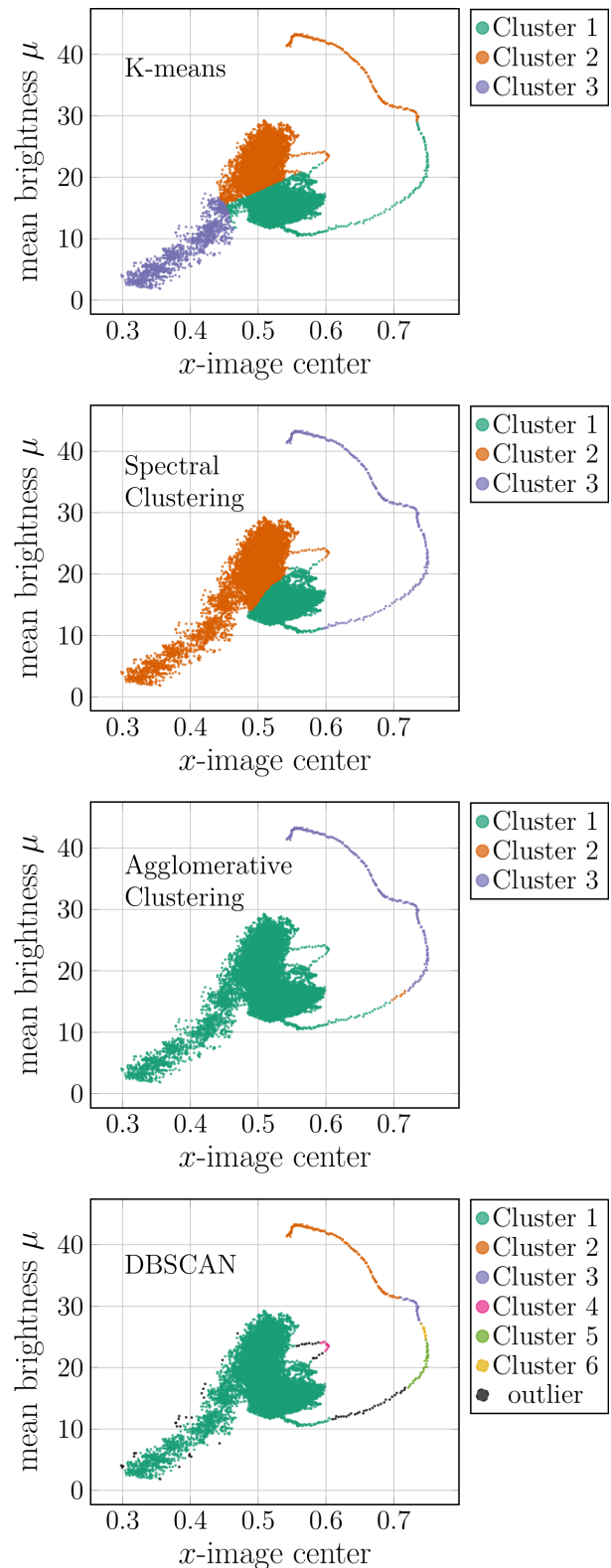


Fig. 4 Comparison of four different clustering algorithms for a low-dimensional data approximation of test 284.

The outcome of DBSCAN at the bottom of Fig. 4 visualizes the result for an intermediate value of ϵ that

identifies six clusters and some outliers. However, more than 95 % of the observations are grouped into the first cluster. This result does not lead to further insights into the dataset.

Apart from DBSCAN, the other three clustering algorithms require the number of clusters K as an input quantity. Here, the best results for the 2D training example and $K = 3$ could be achieved with K-means and with spectral clustering. K-means detects one cluster, colored in purple, that contains the observations in the early phase of the burning test, a further cluster that contains the brightest images with respect to μ from the dataset, and a third cluster that consists of images in which \bar{x} is slightly shifted towards the right side of the image. Compared to K-means, spectral clustering identifies the arc on the right as a separate cluster and returns a nonlinear data separation of the central data point region. On the other hand, the computing time for the spectral clustering was much higher than for K-means, about 70 s instead of 0.8 s, since an additional eigenvalue problem has to be solved then. Using agglomerative clustering, the arc on the right is separated into three clusters but, similar to DBSCAN, most of the data is put into the first cluster which is not useful in our application.

Since K-means clustering led to satisfying results for the 2D test and shows a good scaling behaviour, i.e. it has the lowest computing time compared to the other three approaches, it is chosen as the clustering algorithm used in this study. Nevertheless, since K is not known in advance for the actual dataset, a detailed analysis is still necessary. This analysis is presented in the following section.

4.2 Analysis of the number of clusters

In this section, the number of relevant clusters for each combustion test is determined. For this purpose, all burning tests listed in Table 2 have been clustered with a K-means++ algorithm for different input values for K up to ten. The algorithm was implemented vectorized in Matlab and was executed sequentially on a Linux workstation with 128 GB main memory and an Intel Xeon E5-2697A v4 CPU with 16 cores and 2.6 GHz per core. The maximum memory bandwidth of the system is 76.8 GB/s. A complete clustering of a single test with an image data size of about 8 GB, stored as 8-bit integer values, required about 1.5 days on the workstation. Since the combustion data was converted from integer to floating point numbers to allow a computation of the cluster centroids (arithmetic mean of all images in the cluster), the main memory requirement of the program was in the order of 60 GB. Furthermore, it is noted that

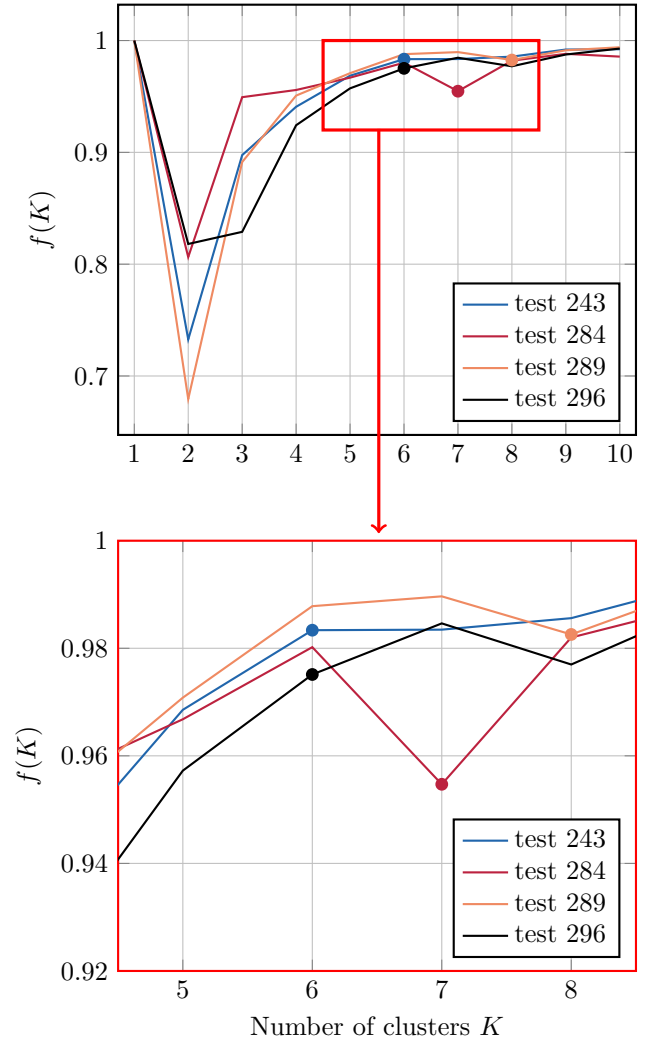


Fig. 5 Evaluation function $f(K)$ for each combustion test to determine the number of clusters K . The choices for K that will be used afterwards is indicated with a dot for each test.

the complexity of K-means type clustering increases linearly with K which restricts the maximum value for K in practice.

As a side note, it is obviously clear that the computing time of the clustering could be reduced by running the algorithm in parallel on all 16 cores of the workstation. Since the algorithm can be parallelized efficiently, a close to optimal parallel scaling behavior is expected. However, the Parallel Computing Toolbox that was used for parallelization in Matlab generates separate copies of the involved matrices on each worker. Then, the memory requirement of the program becomes larger than 60 GB and the workstation runs out of memory in most cases. As a result, an alternative parallel implementation of the algorithm in the conclusion is addressed in Section 5 and, as mentioned before, the program is executed sequentially in the following.

Since a single run of K-means might lead to local optimal minimum solutions for J , all runs of the algorithm for all values of K have been repeated ten times. In all cases, the relative differences in $\min(J)$ in the different runs was below 0.5% and the obtained clusters were almost identical. Furthermore, test 284 was fully clustered a second time with a K-means implementation in the Python-based HPC library HeAT (Krajsek et al. 2018), the Helmholtz Analytics Toolkit, which resulted in almost identical clusters for each K . To the best of our knowledge, this indicates that the clustering results are close to optimal for this specific clustering algorithm.

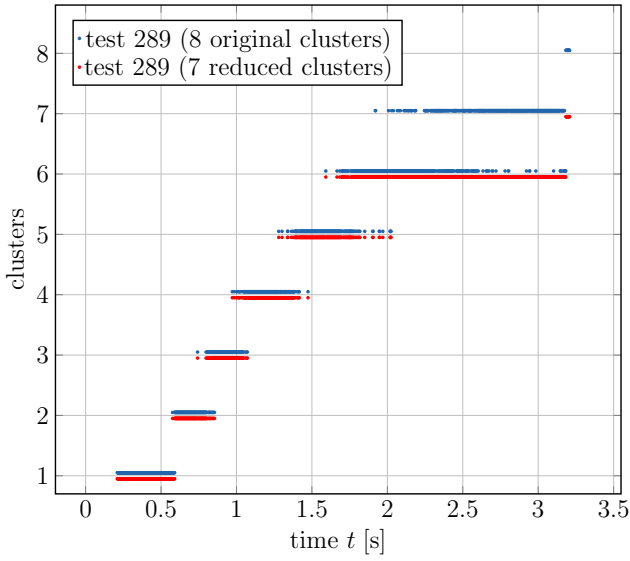


Fig. 6 Distribution of the frames in test 289 for $K = 8$ clusters (blue) and the result of the combination of two overlapping clusters (red).

As explained in Section 2, it is first necessary to determine K for each test. For this purpose, Fig. 5 visualizes $f(K)$ as defined in equations (2) and (3) for all tests. For the values of K that are considered, $f(K)$ always has a global minimum at $K = 2$. This indicates that there are two main structures in the combustion on the coarse scale: an ignition phase and a much longer steady combustion process. But, as it is well known from combustion theory, more structures can be found on a short-time scale. Furthermore, it can be noticed that $f(K = 2) \approx 0.8$ for test 284 but $f(K = 2) \approx 0.7$ for test 289. As mentioned before in Section 2, small values for a specific K indicate that the essential flow characteristics are resolved with this specific choice of K . Therefore, the flow field in test 289 might be better described with $K = 2$ clusters than the flow field in test 284. Since a larger number of clusters seems to be necessary to describe test 284, this might indicate that

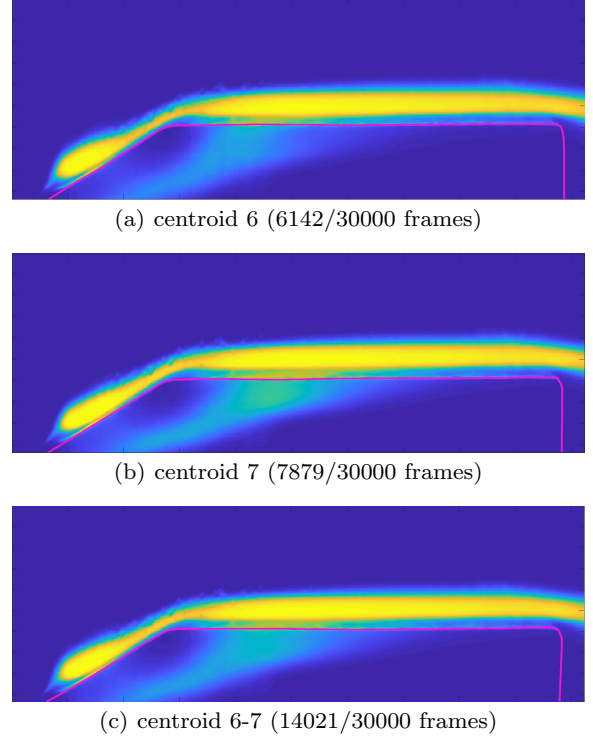


Fig. 7 Since the clusters C_6 and C_7 have a high overlap in time, the corresponding centroids are almost identical. This motivates a recombination of C_6 and C_7 into a single cluster shown in (c). Note that position of the fuel slab is also indicated (pink color).

there are more fluctuations and short-time phenomena in test 284 than in test 289. Next, test 284 has a local minimum $f(K) \approx 0.95$ for $K = 7$. Later on, it is shown that a short-time turbulence is resolved with $K = 7$. Therefore, $K = 7$ clusters are assumed in test 284 and will be interpreted in Subsection 4.3. The outcomes of $f(K)$ in Fig. 5 for the tests 243, 289, and 286 are not conclusive in the first place. In all cases, values for K in the range $2, \dots, 8$ can be justified. In the end, $K = 8$ was chosen for test 289 since, with this number of clusters, the termination phase of the combustion is also resolved. Interestingly, $f(K = 8) \approx 0.98$ is also a local minimum of $f(K)$ for test 289. Finally, the analysis of the tests 243 and 296 in Section 4.3 is performed with $K = 6$ clusters. For these two tests, $K = 6$ is the minimum number of clusters that is required to resolve the extinction phase in both combustion tests. As shown in Fig. 5, $f(K)$ tends to further increase for $K > 6$ so that larger numbers for K cannot be justified using this approach.

For large values of $K > 10$ that were not considered here, further short-time structures might be resolved. In practice, a value of K that is too small has the effect that relevant flow phases are not resolved. On the other

hand, if K is too large the only potential drawback is that identical flow phases, which should be one, are divided in two clusters. But, as the separated structures are connected in time this situation can be identified and corrected. As an illustrative example, the correlation between $K = 8$ clusters and the points in time of the corresponding images for test 289 is plotted in Fig. 6. It can be observed that each cluster belongs to a certain flow phase since they correspond with a certain period in time. However, it is possible to notice that both clusters C_6 and C_7 have a high percentage of overlap in time. A further analysis of the corresponding centroids μ_6 and μ_7 , which can be visualized as images shows that both clusters are almost identical, see Fig. 7. As mentioned before, this is a separation of one flow phase into two almost identical clusters. In order to solve this problem, C_6 and C_7 are recombined. Fig. 6 plots in red the distribution of the frames after a recombination procedure to seven clusters. Now, Fig. 6 only has a minor overlap for the seven clusters in test 289. For that reason, in this study, the recombined dataset with $K = 7$ clusters is used to analyze test 289.

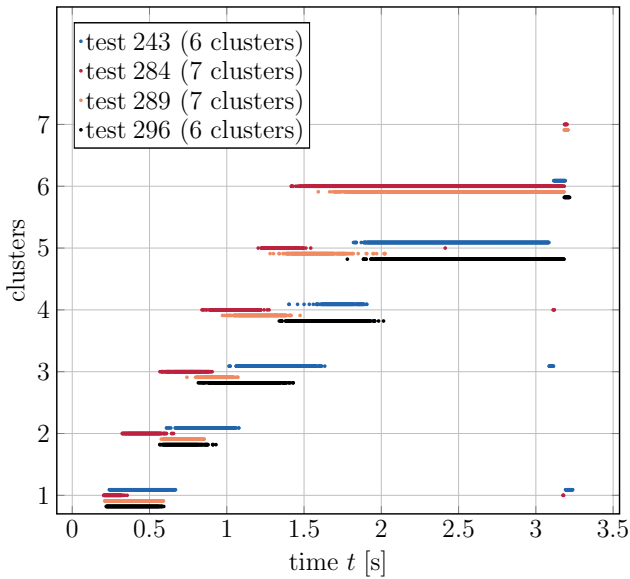


Fig. 8 Classification of all frames in 4 tests to their corresponding clusters. For a better visualization and to avoid an overlap, the points have been slightly shifted in vertical direction.

Finally, Fig. 8 shows the correlation between the clusters and the points in time for all four tests. For a better illustration of the output, the clusters have been ordered in time such that the first cluster represents images from the first flow phase and analogously for all other clusters. Furthermore, the different scatter plots in Fig. 8 have been slightly shifted in vertical

position. The reason for this shift in y -direction is to avoid an overlap of the different tests. Since tests 243 (6 clusters), 284 (7 clusters) and 296 (6 clusters) do not show overlapped clusters in time, a recombination of the clusters as in test 289 is not required. However, such a recombination might be necessary if large values of $K > 10$ are chosen for the clustering algorithm to reveal further short time structures.

As shown in Fig. 8, all tests have a long-time flow phase that contains a high percentage of all frames. In test 243 and 296, this phase has a length of about 1.2 s and belongs to the fifth cluster C_5 . Interestingly, this flow phase is longer in the tests 289 and 284. Here, this phase belongs to the cluster C_6 and has a length of about 1.4 s in test 289 and about 1.7 s in test 284. This result already demonstrates that characteristic experimental features can be identified with a clustering approach. This analysis will be further continued in the following section.

4.3 Analysis of the combustion

The four tests that will be analyzed in the following have been chosen so that the influence of the fuel (tests 284 and 289), oxidizer mass flow (tests 289 and 296) and CH* filter (tests 296 and 243) can be discussed (see Table 2). After each test, the main parameters characterizing the combustion process have been computed, in order to better understand and interpret the optical results. They are reported in Table 3. At this point it is important to underline that the regression rate values reported here are not representative of a typical hybrid engine. Indeed, they are affected by the geometry of the fuel slab (normally a fuel grain is used) as well as by the interaction of the oxidizer flow with the chamber walls, which also reduces the temperature in the combustion chamber, and by side burning effects. Moreover, inaccuracies also arise from the short burning times since the ignition transients have a bigger influence then. The results obtained from the combustion visualizations give many insights into the hybrid combustion process and allow to separate the different burning phases.

Table 3 Test parameters (all values are averaged in space and time).

Test	\dot{m}_{Ox} [g/s]	OF	\dot{r} [mm/s]	G_{Ox} [kg/(m ² s)]
284	46.45	7.5	0.2841	7.5
289	46.5	12.16	0.24	9.2
296	12.43	4.8	0.157	2.42
243	11.8	13.6	0.0545	2.37

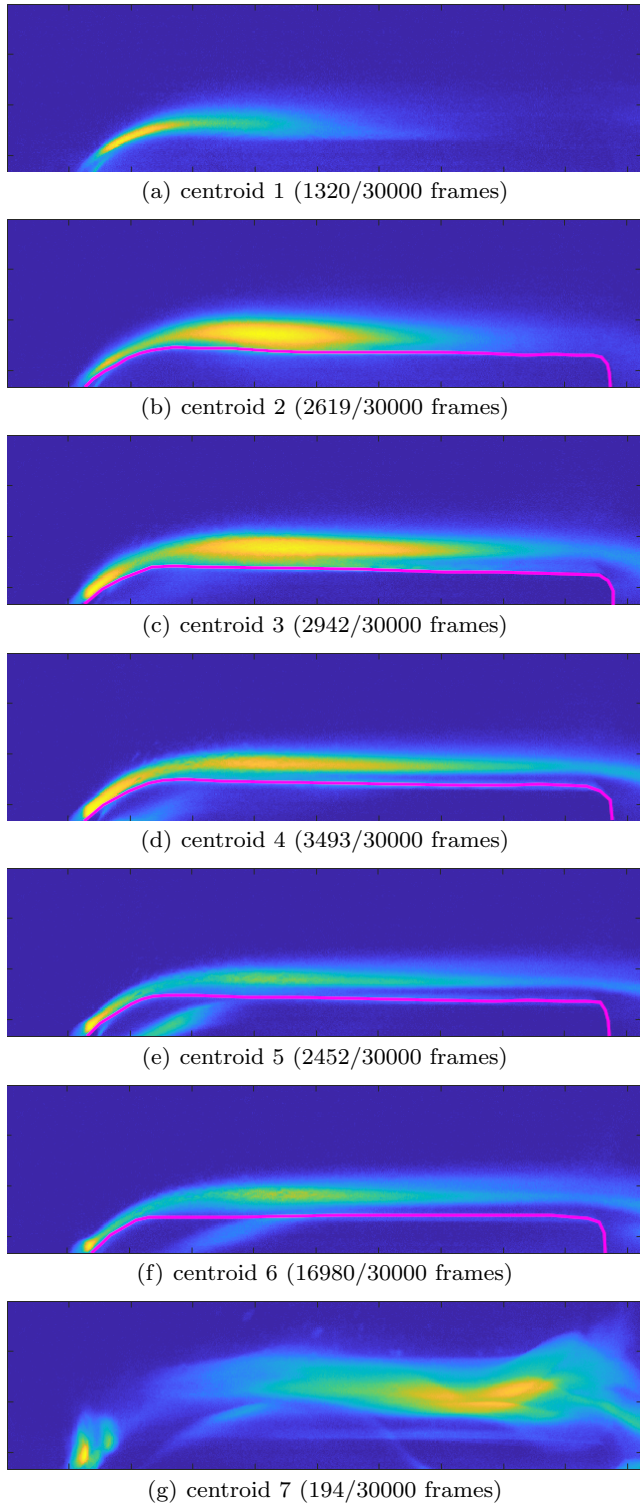


Fig. 9 Visualization of the seven centroids in test 284.

First of all, it is interesting to have a look at test 284. This combustion video was realized with a band-pass filter centered around 431 nm placed in front of the high-speed camera. It is generally recognized (Devriendt et al. 1996; Schefer 1997) that the primary species

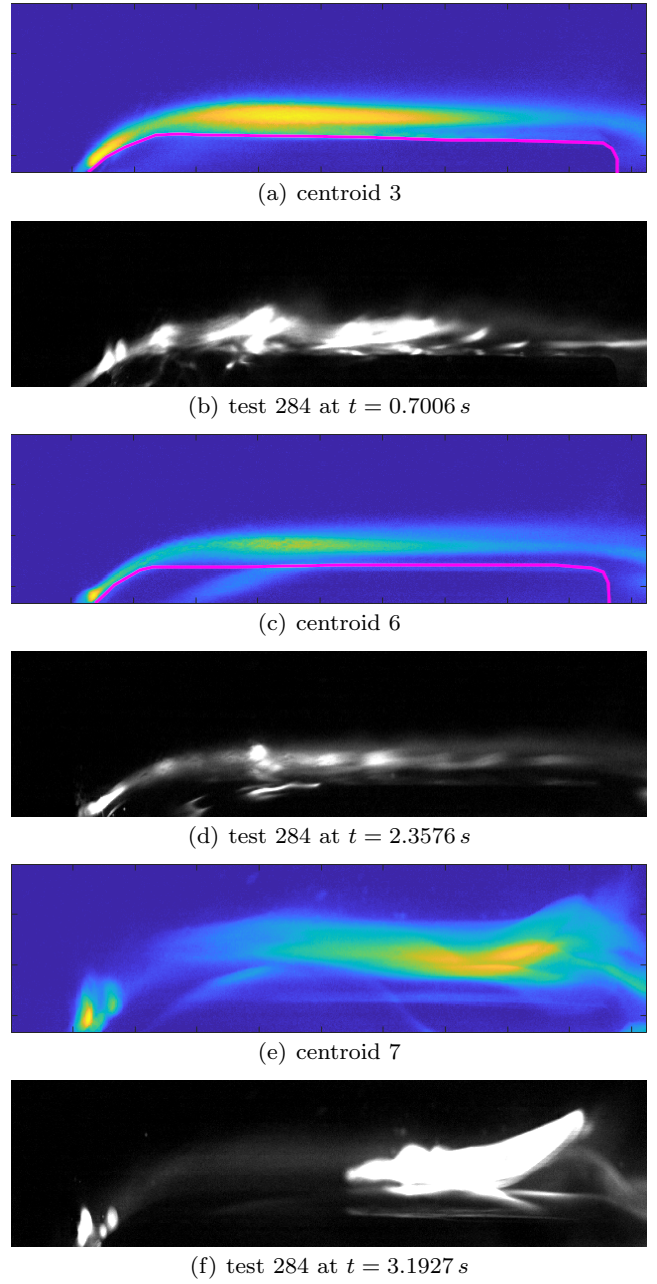


Fig. 10 Comparison between individual frames and the corresponding centroids for test 284.

contributing to flame luminescence are the electronically excited species CH^* , C_2^* and OH^* . All three species show a close correspondence across the main reaction zone and are thus equally suitable as markers for the flame zone location. In particular, the concentrations of CH^* increase rapidly to a maximum within the flame and then decay rapidly downstream of the reaction zone (Schefer 1997). Therefore, the CH^* images of test 284 give a good representation of the main flame location. Looking at Fig. 9, it is possible to notice that 7 different clusters are separated by the algo-

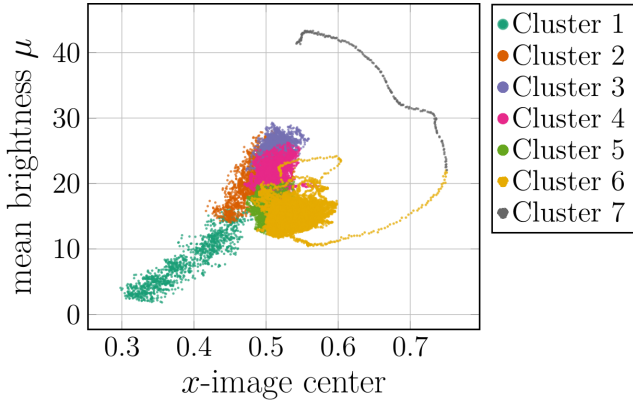


Fig. 11 Low-dimensional data approximation of test 284. The color indicates the corresponding cluster 1, ..., 7 for each image.

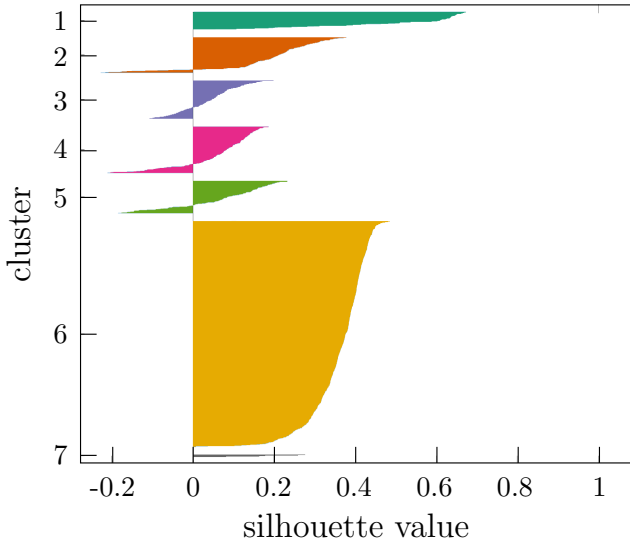


Fig. 12 Silhouette plot of test 284. The plot illustrates the clustering quality of each image to one of the $K = 7$ clusters.

rithm. If possible, the position of the fuel slab (colored in pink) has been added to simplify an interpretation of the centroids. The first cluster goes from $t = 0.2$ s until $t = 0.33$ s and represents the initial ignition phase. In fact, until $t = 0.3$ s the valves of the torch igniter are open (see Table 1) and the gases coming from the hydrogen/oxygen combustion enter the combustion chamber from the bottom head of the fuel slab. They start gasify the fuel, which, in turn, initiates burning with the oxygen entering the chamber. Since the ignition is coming from the bottom of the combustion chamber, the first area that starts to gasify and to burn is exactly the fuel step, as it can be seen in the image of centroid 1 in Fig. 9. Clusters 2, 3 and 4 are representing the different phases of the fuel ignition transient and are going from $t = 0.33$ s until around $t = 1.23$ s. Cluster 2 is starting directly after the closing of the ignition valves,

thus no combustion gases are coming from the igniter any more and the fuel slab starts to burn without any energy addition from the outside. This is the start of a self-sustained combustion. At this point, the oxygen mass flow is not yet settled down ($\dot{m}_{O_2} < 10$ g/s), but it is still increasing; therefore, there is still no steady-state. As the fuel slab starts to burn, different phases can be observed in these 3 clusters: at the beginning (cluster 2) the fuel slab is burning just in the front, then, as the time goes on (and the oxygen mass flow increases $\dot{m}_{O_2} < 25$ g/s), it starts to burn in the middle (cluster 3, Fig. 10 (a)-(b)) and, finally, in cluster 4 (\dot{m}_{O_2} reaches 40 g/s), almost the whole surface is burning. It is also possible to notice that, in cluster 4, the brightness of the flame already starts going down. This means that the temperature in the combustion chamber is already quite high and a quasi steady-state flame is appearing. The result might be unexpected at first, but it is important to remember that the CH^* radicals are more likely to appear at the beginning of the combustion process, when the temperature is not yet too high. Later, they are further reacting (the most likely reaction is: $CH^* + O_2 = CO + OH^*$) and, therefore, their concentration decreases while the OH^* concentration increases. So, it is more likely that the brighter burning regions are seen at the beginning of the combustion (clusters 2,3,4), while during the state-state phase (clusters 5 and 6), the concentration of the CH^* decreases and so does also the flame brightness (see Fig. 10). Clusters 5 and 6 also show a more or less constant flame thickness on the whole fuel slab length. In the time window of these two clusters (1.23 s $< t < 3.18$ s), the oxidizer mass flow is more or less steady (it increases from 45 to 50 g/s), the temperature reached the maximum flame temperature and the flame is well settled down on the fuel slab surface. This situation does not change until the closing of the oxygen main valve and the opening of the nitrogen purging at $t = 3.18$ s, when cluster 7 starts. This last phase represents the combustion extinction phase. In the video, it is possible to see that the nitrogen purge comes at around $t = 3.08$ s until $t = 3.1$ s. At $t = 3.18$ s the flame is completely extinguished. This causes, in turn, a drop in the chamber pressure, which practically causes a suction of still-burning paraffin gases from the rear part of the fuel slab. In the video it is clearly visible that these reacting gases are coming from the back and they are travelling until the fuel slab head (see Fig. 10). So, cluster 7 represents the flame extinguish phase, where the remaining paraffin gases are still burning with the rest of the oxygen available in the combustion chamber.

As it can be noticed from Fig. 11 and 12, only cluster 1 (first ignition transient), 6 (steady state) and 7 (extinc-

tion) are well separated. The clusters 2,3,4 and 5 contain elements that could also belong to other clusters. This happens because the transition between the ignition transient and the steady state phase is not well defined. The flow dynamics and, consequently, the flame brightness change slowly and gradually and it is not really possible to define when one phase is finished and a new one starts. Thus, these four clusters represent the transition between the ignition and the steady state phases. On the other hand, the transition between the steady state and the extinction phase is well defined because it begins exactly when the nitrogen purge starts (at $t = 3$ s). This moment can be easily recognized in the small arch belonging to cluster 6 in Fig. 11. When the nitrogen enters the combustion chamber, the mean brightness first increases and moves to the front of the fuel and then decreases and moves more to the back. Finally, during the extinction phase (cluster 7), the brightness increases and moves again to the front of the fuel (due to the "suction" effect), as it can be seen in Fig. 11. From this figure, it can be also noticed that, except for the ignition and the extinction phases, the combustion is mainly taking place in the middle of the fuel slab, where the center of the mean brightness is placed.

Having a look at the results of test 289 and comparing them with those of test 284, two main observations can be done: both tests present the same number of clusters (see Fig. 8); the mean flame brightness of test 289 is higher than that of test 284 and it is more shifted to the back of the fuel slab (see Fig. 13 and 14). The presence of the same number of clusters is due to the fact that, having in both tests the same oxidizer mass flow, the flow dynamics is similar and, thus, the different combustion phases take place more or less at the same point in time (small differences are due to the fact that fuel 6805+5% polymer needs a bit longer to ignite). In fact, also in test 289, the first 3 clusters ($0.2\text{ s} < t < 1\text{ s}$) represent the ignition transient (\dot{m}_{Ox} reaches 30 g/s at the end of cluster 3); cluster 4 and 5 ($1\text{ s} < t < 1.8\text{ s}$) show the transition from the ignition to the steady-state (\dot{m}_{Ox} reaches 45 g/s); cluster 6 ($1.8\text{ s} < t < 3.18\text{ s}$) represents the steady state, where the oxidizer mass flow reaches 50 g/s; finally, cluster 7 ($3.18\text{ s} < t < 3.2\text{ s}$) shows the extinction phase. On the other hand, since the fuels burnt in the two tests have different regression rates (the pure paraffin 6805 burns faster than the blend), their combustions are characterized from different oxidizer-to-fuel mixture ratio (OF) and, thus, different flame temperature and brightness. In particular, since the OF of the blended fuel is higher than that of the pure paraffin, the flame temperature in test 289 is lower than in test 284. Keeping in mind

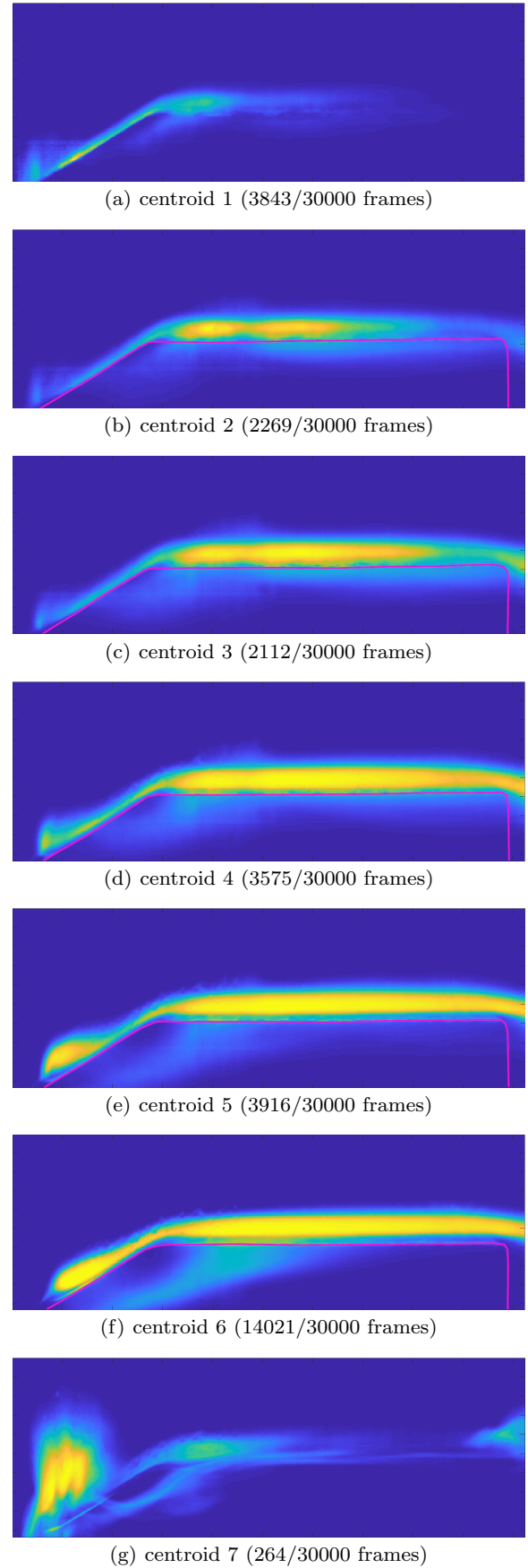


Fig. 13 Visualization of the centroids in test 289. Note that the original centroids 6 and 7 have been combined into a combined centroid 6.

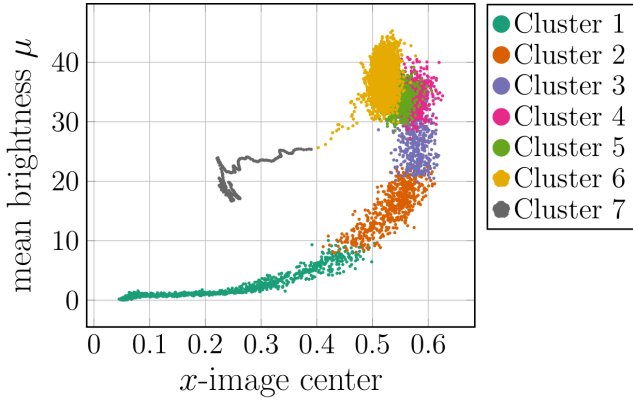
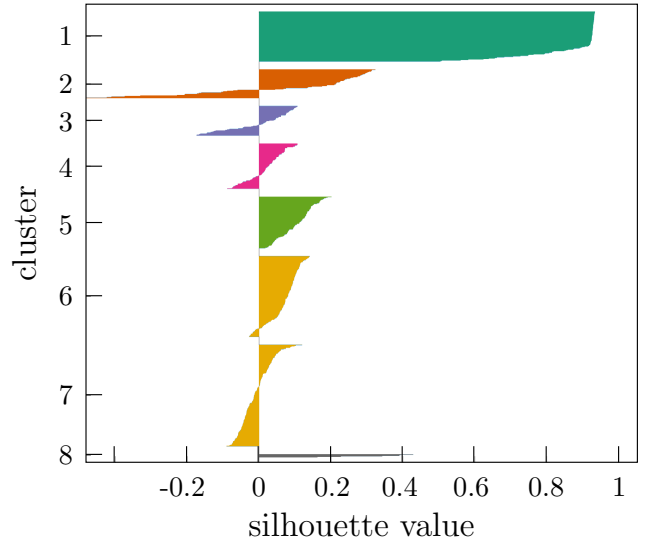


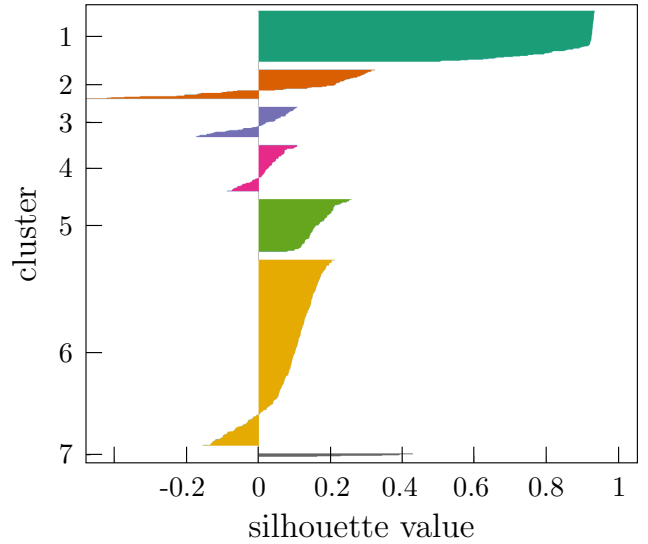
Fig. 14 Low-dimensional data approximation of test 289 with color indication for $K = 7$ clusters. Note that the clustering was originally performed for eight clusters and that the two clusters 6 and 7 have been recombined.

that also test 289 is performed with a CH^* filter, this means that the CH^* concentration is higher in test 289 and thus also the flame brightness. In this case, clusters 1 (first ignition transient), 5 (transition to the steady state) and 7 (extinction) are well separated, as it can be seen in Fig. 15 (b). Cluster 6, which represents the steady state phase, has some negative values that probably belong to other clusters. However, compared to the original output of the clustering algorithm shown in Fig. 15 (a), the recombination of C_6 and C_7 into a new cluster C'_6 (cf. Sec. 4.2 for a detailed explanation) reduced the number of images with negative silhouette values. Furthermore, Fig. 15 (a) justifies the performed cluster recombination since the original cluster 7 has more negative than positive silhouette values. From our point of view, such a situation, i.e. a cluster with more negative than positive silhouette values, is a good indicator for a potential recombination with its neighboring clusters in the case that a time-dependent dataset with a high temporal resolution is considered. As mentioned before in the description of Fig. 3, due to the high temporal resolution of the camera, the difference between adjacent images is relatively small. Therefore, a misclassification of images in the dataset, indicated with negative silhouette values, should primarily occur between adjacent clusters. On the other hand, this does not apply to general datasets, i.e. in situations in which there is no dense temporal sampling of the data points. Then, misclassified images could belong to any other cluster.

In order to study the influence of the oxidizer mass flow, it is necessary to analyze the results of test 296 and compare them with test 289. First of all, it can be noticed that a lower number of clusters is found in test 296 by the clustering algorithm. In fact, since the oxidizer mass flow of test 296 is lower than that of test



(a) original clustering result



(b) after recombination of C_6 and C_7

Fig. 15 Silhouette plot of test 289. The plot illustrates the clustering quality for the original clustering result (a) and after the recombination of C_6 and C_7 into a combined cluster C'_6 (b).

289 (10 g/s instead of 50 g/s), the flow dynamics is slower, the turbulence level is lower and the transient presents less variability with respect to test 289. In this case, cluster 1 and 2 ($0.2\text{ s} < t < 0.9\text{ s}$) represent the ignition phase (\dot{m}_{Ox} reaches about 6 g/s at the end of cluster 2); cluster 3 and 4 ($0.9\text{ s} < t < 1.9\text{ s}$) show the transition from the ignition to the steady-state (\dot{m}_{Ox} reaches 10 g/s); cluster 5 ($1.2\text{ s} < t < 3.18\text{ s}$) represents the steady-state, with a constant oxidizer mass flow of about 13 g/s; finally, cluster 6 ($3.18\text{ s} < t < 3.2\text{ s}$) represents the extinction phase (see Fig. 16). Moreover, looking at Fig. 17, it can be noticed that the mean

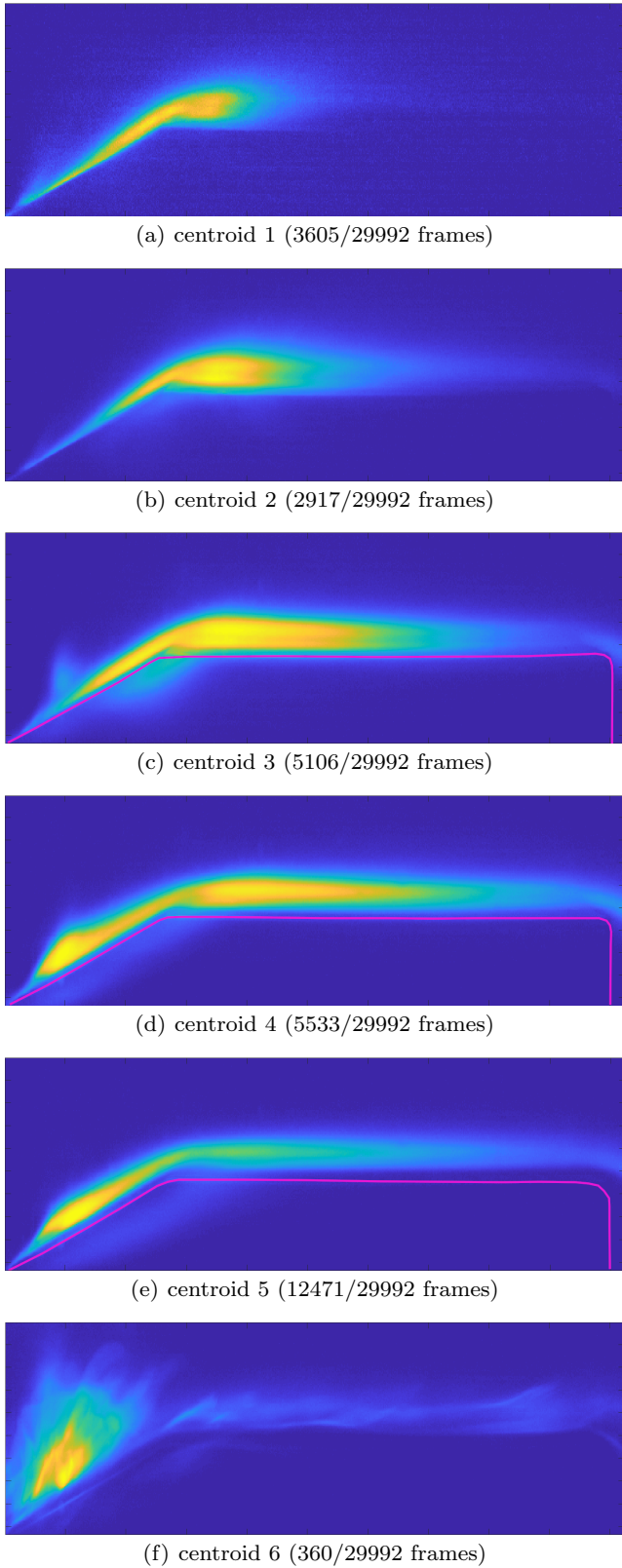


Fig. 16 Visualization of the six centroids in test 296.

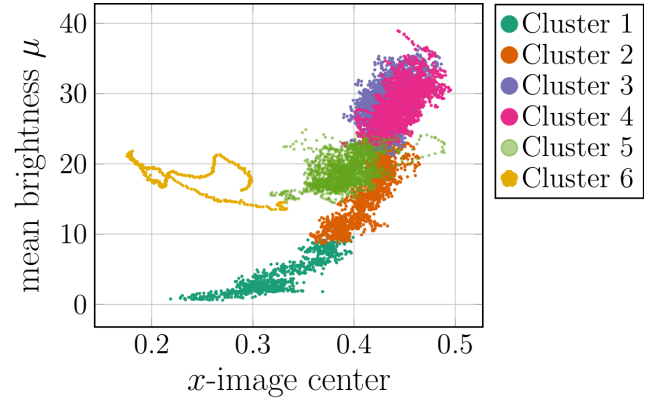


Fig. 17 Low-dimensional data approximation of test 296 with color indication for $K = 6$ clusters.

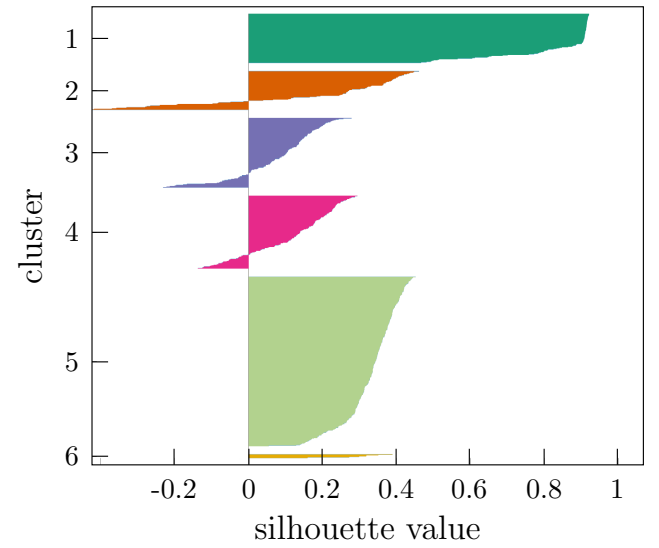


Fig. 18 Silhouette plot of test 296. The plot illustrates the clustering quality of each image to one of the $K = 6$ clusters.

flame brightness is a bit lower and has a higher variance than that of test 289. This is due to the fact that, with a lower oxidizer mass flow, the OF of test 296 is lower than that of test 289 and thus the flame temperature is higher. As previously discussed, at higher temperature the CH^* concentration decreases in favor of the OH^* concentration, leading to a lower brightness of the CH^* emission. Finally, in test 296, the center of the mean flame brightness is moved more to the front of the fuel slab (see Fig. 17). Also in this case, the silhouette plot in Fig. 18 shows that only the clusters 1, 5 and 6 are precisely separated from the others. They are representing the three main combustion phases: ignition, steady-state and extinction. On the other hand, clusters 2,3 and 4 represent the transition between the ignition and the steady-state and each of them contains elements that could also belong to other clusters. As already expected, this kind of noise (the negative val-

ues in the silhouette plot) is decreasing while approaching the steady-state phase, meaning that the brightness variability between the frames is getting smaller and the algorithm is able to determine an unambiguous cluster.

Concerning the same fuel configuration but without filter (test 243), it exhibits the same number of clusters and the same phases as test 296 (same oxidizer mass flow), see Fig. 19. The first 3 clusters ($0.2\text{ s} < t < 1.6\text{ s}$) represent the ignition transient of the fuel slab (note that the first visible flame, which corresponds to the first self-sustained flame, is only visible starting at $t = 0.35\text{ s}$). The first cluster represents the initial flame, which is not yet developed over the whole fuel slab length. It seems like a recirculation zone where most likely the paraffin gases are entrained in a kind of vortex (probably caused by the step) and start burning with the oxidizer. Then, the vortex becomes bigger (cluster 2) and more paraffin gases get entrained in the recirculation area, until the whole fuel slab burns (cluster 3). It is possible to note that in test 243 there is always a small recirculation area directly after the step, which is not burning. The oxidizer is not able to reach that area because it burns completely before and therefore no combustion is occurring there. Cluster 4 ($1.6\text{ s} < t < 1.95\text{ s}$) represents the transition between the ignition and the steady-state, while cluster 5 ($1.95\text{ s} < t < 3.16\text{ s}$) represents the steady-state. As it can be noticed from Fig. 19, the difference between both clusters is small. This is again due to the fact that there is no clear distinction between the end of the ignition phase and the beginning of the steady-state. The steady-state flame without the filter is a bit brighter than the one given by the CH^* emission light, at the same operating conditions; but the center of the mean brightness is placed, also in this case, more to the front of the fuel slab (see Fig. 20). Finally, cluster 6 ($3.16\text{ s} < t < 3.2\text{ s}$) represents the extinction phase. During this phase, the flame becomes brighter and moves more to the back of the fuel slab (see Fig. 20). As can be observed, the essential differences between the tests with and without the filter are the transients. In fact, the initial ignition phase, when the ignition valve is still open ($0.2\text{ s} < t < 0.33\text{--}0.35\text{ s}$), is not visible in the test without the filter (nothing is visible in the video of tests 243 before $t = 0.35\text{ s}$). The brightness of the combustion gases of the torch igniter is too weak to be visible without the filter. Thus, the first visible flame is that of the fuel slab itself when it starts burning. Also the extinction phase of test 243 is different from that of test 296. This is, of course, also due to the filter that allows us to see only the CH^* emissions. Again, the ignition, steady state and extinction phases are well separated from the others, as it can be noticed from Fig. 21. The other clusters contain

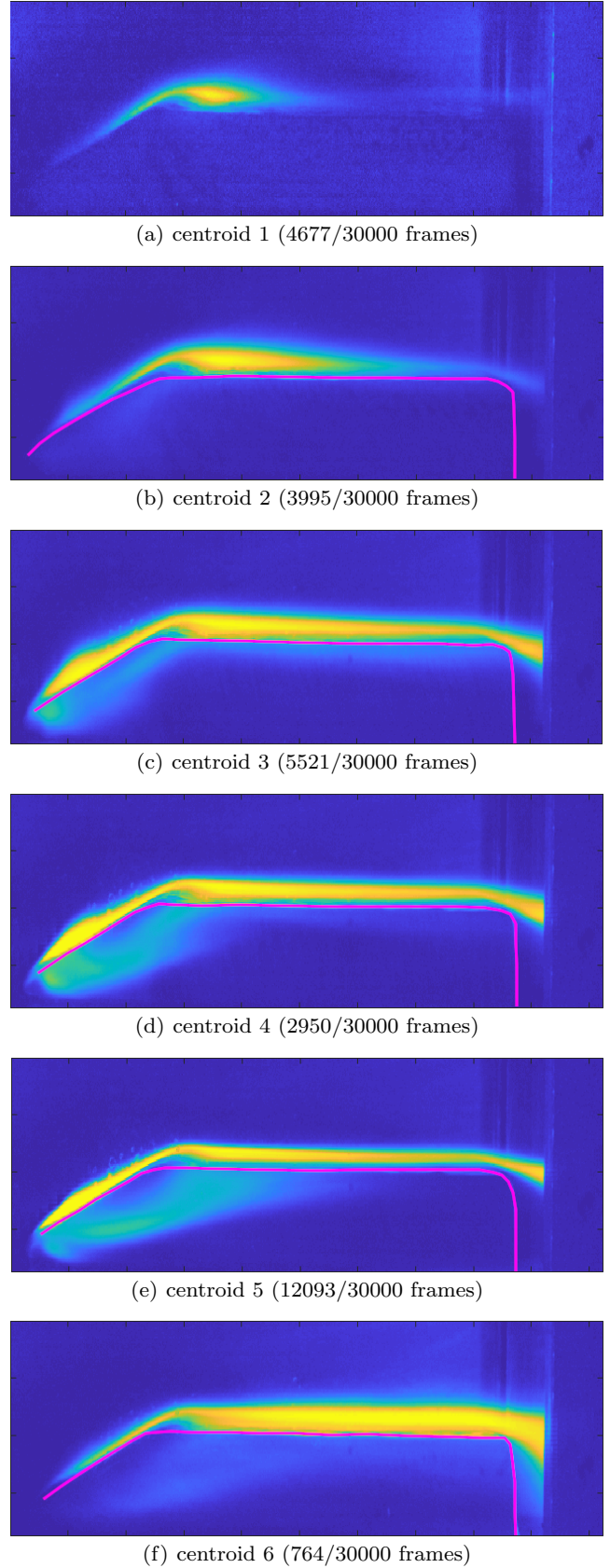


Fig. 19 Visualization of the six centroids in test 243.

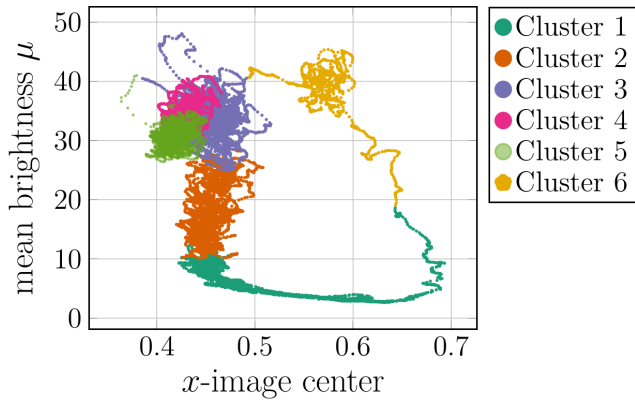


Fig. 20 Low-dimensional data approximation of test 243 with color indication for $K = 5$ clusters.

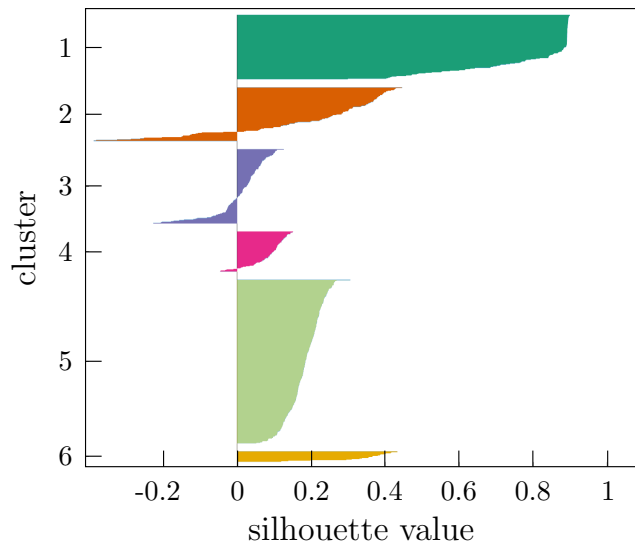


Fig. 21 Silhouette plot of test 243.

elements that do not really belong to a precise cluster since they represent the transitions phases.

Table 4 Time length of each cluster [s].

Test	C_1	C_2	C_3	C_4	C_5	C_6	C_7
243	0.47	0.4	0.55	0.3	1.21	0.08	x
284	0.13	0.26	0.29	0.35	0.25	1.7	0.02
289	0.38	0.23	0.21	0.36	0.39	1.4	0.03
296	0.36	0.29	0.51	0.55	1.25	0.04	x

Finally, the output of the clustering algorithm is used for a quantitative comparison of the four tests. For this purpose, Table 4 lists the time length of each cluster. The table summarizes the number of images in the different clusters (cf. Fig. 9, Fig. 13, Fig. 16 and Fig. 19 for the precise number of the images). It is directly clear from Fig. 8 that some clusters overlap in

time and that some images are misclassified. Therefore, Table 4 shows the output of the pure clustering without any postprocessing by the user. Nevertheless, Table 4 allows to compare the different test configurations. First, the steady state phase of the combustion (C_5 in test no. 243 and 296, C_6 in test no. 284 and 289) has been written bold to simplify a comparison. Since the tests no. 284 and 289 have a longer steady combustion, these tests perform better in the case that the length of the steady combustion phase is used as a quality measure. The small difference in the time length of C_6 is due to the fact that different fuel compositions are used in the two tests, thus having slightly different ignition delays. On the other hand, the steady combustion phases in the tests no. 243 and 296 have the same time length, since the two tests are characterized by the same fuel composition and oxidizer mass flow. Second, the length of the ignition phase resolved in the first cluster can be considered as a further quality measure and gives information about the ignition delays. Here, it is noted that test 284 has a considerably shorter first flow phase than the other tests. It has to be noted, however, that slightly different flow field conditions are captured in C_1 which complicates a direct comparison of the tests. Therefore, a proper matching of the flow field conditions in the different clusters is essential to allow for a more detailed analysis of the different test configurations and a direct quality measure for comparison.

5 Conclusion

A hybrid combustion dataset was structured and analyzed with K-means++ clustering. To the best of our knowledge, this is the first application of image clustering in the specific application of hybrid rocket combustion. The analysis revealed several interesting short-time phenomena in the dataset and clearly indicate the potential of unsupervised learning techniques for the structuring of large datasets. In this study, it was proved that the clustering algorithm is able to recognize the different burning phases in the analyzed tests. A clear distinction between ignition, steady-state and extinction phases was performed and a valuable insight into the combustion dynamics (especially of the transients) was obtained. In particular, differences in the number of clusters were observed depending on the oxidizer mass flow and, thus, on the flow dynamics. A lower oxidizer mass flow corresponds with a shorter ignition transient and, consequently, with a lower number of clusters. Moreover, differences, especially in the transients and in the flame brightness, were also noticed between tests with and without the CH^* filter. In particular, the steady-state flame is less bright in the tests with

the filter, due to the fact that the CH^* radicals are further reacting at high temperatures, thus reducing their concentration. As a further result, a test evaluation that bases on the detected clusters allows to quickly determine the performance of a specific configuration (cf. Table 4). The analysis that has been presented here becomes especially important if a large number of different tests is performed. Then, the clustering allows to identify an optimal configuration in the dataset.

Furthermore, spectral clustering, a computationally more intensive approach shown in Fig. 4 might resolve further short-time structures and give further insights into the complex combustion process. A parallel implementation of this algorithm in the HPC library HeAT (Krajsek et al. 2018) could lead to enormous reductions in computing time such that it could for the analysis of the full dataset.

The intention of this work is to demonstrate that unsupervised learning techniques can be successfully applied to large combustion datasets. In the future, the authors will further develop unsupervised, as well as supervised, learning techniques to apply to combustion data. The objective is to get detailed information on the important phenomena characterizing the hybrids combustion process, such as the entrainment. The next step will be to train the algorithm how to automatically recognize particular structures, such as Kelvin-Helmholtz waves, vortices and droplets, in space and time.

Acknowledgements This research was carried out under the project *Antriebstechnologien und Komponenten für Trägersysteme* (ATEK) by the German Aerospace Center (DLR).

References

- Arthur D, Vassilvitskii S (2007) K-means++: The advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, SODA 2007, pp 1027–1035, URL <http://dl.acm.org/citation.cfm?id=1283383.1283494>
- Ciezki HK, Sender J, Clauß W, Feinauer A, Thumann A (2003) Combustion of solid-fuel slabs containing boron particles in step combustor. *J Propul Power* 19(6):1180–1191, DOI 10.2514/2.6938
- Devriendt K, Hook HV, Ceusters B, Petters J (1996) Kinetics of formation of chemiluminescent CH by the elementary reactions of C_2H with O and O_2 : a pulse laser photolysis study. *Chem Phys Lett* 261:450–456
- Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, AAAI Press, KDD'96, pp 226–231, URL <http://dl.acm.org/citation.cfm?id=3001460.3001507>
- Hastie T, Tibshirani R, Friedman J (2009) Hierarchical clustering. *The Elements of Statistical Learning* 2
- Jain A (2010) Data clustering: 50 years beyond k-means. *Pattern Recogn Lett* 31(8):651–666
- Karabeyoglu A, Cantwell B, Altman D (2001) Development and testing of paraffin-based hybrid rocket fuels. In: 37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, American Institute of Aeronautics and Astronautics, Salt Lake City, Utah, DOI 10.2514/6.2001-4503
- Karabeyoglu A, Altman D, Cantwell BJ (2002) Combustion of liquefying hybrid propellants: Part 1, general theory. *J Propul Power* Vol.18(No. 3):610–620, DOI 10.2514/2.5975
- Karabeyoglu A, Stevens J, Geyzel D, Cantwell B, Micheletti D (2011) High performance hybrid upper stage motor. In: 47th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, American Institute of Aeronautics and Astronautics, DOI 10.2514/6.2011-6025
- Kobald M, Petrarolo A, Schlechtriem S (2015) Combustion visualization and characterization of liquefying hybrid rocket fuels. In: 51st AIAA/SAE/ASEE Joint Propulsion Conference, American Institute of Aeronautics and Astronautics, DOI 10.2514/6.2015-4137
- Krajsek K, Comito C, Götz M, Hagemeyer B, Knechtges P, Siggel M (2018) The Helmholtz analytics toolkit (heat): a scientific big data library for hpc. In: Extreme Data Workshop 2018, URL <https://elib.dlr.de/124422/>
- Lloyd S (1982) Least squares quantization in pcm. *IEEE T Inform Theory* 28(2):129–137
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley, Calif., pp 281–297, URL <https://projecteuclid.org/euclid.bsm/1200512992>
- Ng AY, Jordan MI, Weiss Y (2001) On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems, MIT Press, pp 849–856
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Petrarolo A, Kobald M (2016) Evaluation techniques for optical analysis of hybrid rocket propulsion. *J Fluid Sci Technol* 11(4):JFST0028–JFST0028, DOI 10.1299/jfst.2016jfst0028
- Petrarolo A, Kobald M, Schlechtriem S (2018) Understanding Kelvin-Helmholtz instability in paraffin-based hybrid rocket fuels. *Exp Fluids* p 59:62, DOI 10.1007/s00348-018-2516-1
- Pham D, Dimov S, Nguyen C (2005) Selection of k in k-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: J Mech Eng Sci* 219(1):103–119
- Rousseeuw P (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
- Schefer RW (1997) Flame sheet imaging using CH chemiluminescence. *Combust Sci Technol* 126(1-6):255–279, DOI 10.1080/00102209708935676
- Sculley D (2010) Web-scale k-means clustering. In: Proceedings of the 19th international conference on World wide web, ACM, pp 1177–1178
- Thumann A, Ciezki HK (2002) Combustion of energetic materials, chap. Comparison of PIV and Colour-Schlieren measurements of the combustion process of boron par-

- ticle containing soild fuel slabs in a rearward facing step combustor, vol 5. Begell House Inc., DOI 10.1615/IntJEnergeticMaterialsChemProp.v5.i1-6.770
- Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. J R Statist Soc B 63(2):411–423